BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C07H 21/04, C07K 14/47, C12N 5/06, 5/08, C12Q 1/68 — A1

(11) International Publication Number: WO 97/38004

(43) International Publication Date: 16 October 1997 (16.10.97)

(21) International Application Number: PCT/US97/05903

(22) International Filing Date: 10 April 1997 (10.04.97)

(30) Priority Data:

08/630,592	10 April 1996 (10.04.96)	US
08/701,380	22 August 1996 (22.08.96)	US
08/706,292	4 September 1996 (04.09.96)	US
08/714,991	17 September 1996 (17.09.96)	US

(71) Applicants: SEQUANA THERAPEUTICS, INC. [US/US]; Suite 160, 11099 North Torrey Pines Road, La Jolla, CA 92037 (US). THE JACKSON LABORATORY [US/US]; 600 Main Street, Bar Harbor, ME 04609 (US).

(72) Inventors: NISHINA, Patsy; The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 (US). NOBEN-TRAUTH, Konrad; The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 (US). NAGGERT, Juergen; The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609 (US). NORTH, Michael; Sequana Therapeutics, Inc., Suite 160, 11099 North Torrey Pines Road, La Jolla, CA 92037 (US).

(74) Agents: BOZICEVIC, Karl et al.; Bozicevic & Reed L.L.P., Suite 200, 285 Hamilton Avenue, Palo Alto, CA 94301 (US).

(81) Designated States: AU, CA, JP, MX, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.

Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: GENE FAMILY ASSOCIATED WITH NEUROSENSORY DEFECTS

(57) Abstract

Nucleic acid compositions are provided that encode a family of mammalian proteins expressed in the retina and brain. Members of the gene family are genetically linked to various neurosensory defects, including cochlear degeneration, peripheral retinal degeneration and cone-rod retinal dystrophy. The nucleic acid compositions find use in identifying DNA sequences encoding homologous or related proteins; for production of the encoded protein; and in studying associated physiological pathways. In addition, modulation of the gene activity *in vivo* is used for prophylactic and therapeutic purposes, such as treatment of neurosensory defects, identification of retinal cells based on expression, and the like. The DNA is further used as a diagnostic for genetic predisposition to the linked neurosensory defect.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

GENE FAMILY ASSOCIATED WITH NEUROSENSORY DEFECTS

Introduction

Sensory neurons give us our perception of the world, by transducing phenomena such as light and sound into signals that can be received and understood by the brain. However, neurons can also be fragile, and susceptible to a number of hereditary and/or age related degenerative disorders. Understanding the genes and gene products that comprise and control neurosensory signaling pathways may provide the basis for future medical advances in this area.

Neurodegenerative disorders result from the premature death of nerve cells in the brain and spinal cord; for example tracts of the acoustic system in degenerative hearing disorders. Such neuronal degeneration has been attributed to genetic defects, transmissible infectious agents, toxic substances, immune system disorders and other as yet undetermined mechanisms. A recent hypothesis is that active photoreceptor cell death, which is characteristic of these genetically distinct disorders, is mediated by a common induction of apoptosis.

Inherited eye disorders are the major cause of childhood blindness in the developed world. Many of these are retinal dystrophies. The retina is the sensory tunic of the eye, containing light sensitive receptors, a complex of neurons, and pigmented epithelium, arranged in discrete layers. In humans, the macula is the portion of the retina that lies directly behind the lens. Cones, the photoreceptor cells responsible for central vision, are heavily concentrated in the macula. The peripheral retina is composed mainly of rods, which are responsible for side and night vision.

Choroidoretinal dystrophies and degenerations, all of which are currently incurable and untreatable, are a common form of retinal dystrophy. Cone-rod retinal dystrophy (CRD) is a severe example, characteristically leading to early blindness. A loss of color vision and visual acuity is accompanied by widespread, advancing retinal pigmentation and chorioretinal atrophy of the central and peripheral retina. Linkage analysis of a large lineage of autosomal dominant CRD has mapped the disease to chromosome 19q, linked to the polymorphic marker *D19S47*. It has been suggested that the disease locus for CRD,

which affects the central as well as peripheral retina, may also be involved in age-related macular degeneration (ARMD).

Hereditary peripheral retinopathies are also relatively common. Retinitis pigmentosa (RP), for example, affects approximately 1.5 million people worldwide. Substantial genetic heterogeneity has been observed in this condition, with over 20 chromosomal loci identified. A predisposition to retinitis pigmentosa can be inherited by autosomal dominant, autosomal recessive, X-linked or digenic modes. In spite of causal heterogeneity, there is significant clinical similarity among RP subtypes. Common signs and symptoms include early electroretinographic abnormalities, ophthalmoscopic findings, and progressively worsening tunnel vision.

It is interesting to note that the mouse mutation, *tubby*, leads to both retinal and cochlear degeneration, indicating a common element in both sensory pathways. It has also been observed that rare monogenic forms of human severe obesity are often accompanied by blindness and deafness: the best characterized are Bardet Biedl syndrome and Alstrom syndrome. Studying these diseases, although important in their own right, may also provide critical clues to the molecular mechanisms leading to an obese state.

The prevalence and clinical consequences of sensory neuronal defects make it of interest to characterize *tubby* and related genes that may be associated with vision and hearing defects.

Relevant Literature

Overviews of photoreceptor dystrophies may be found in Cotlier *et al.* (1995) Surv. Ophthalmology **40**:51-61; Bird (1995) Am. J. Ophthal. **119**:543-562; and Adler (1996) Arch Ophthal. **114**:79-83. Evans *et al.* (1994) Nature Genetics **6**:210-213 describes the genetic mapping of cone-rod retinal dystrophy. Shugart *et al.* (1995) Am J Hum Genet. **57**:499-502 disclose fine genetic mapping of a gene for autosomal recessive retinitis pigmentosa (RP 14) on chromosome 6p21. Berson (1996) Proc Natl Acad Sci USA **93**:4526-4528 review retinitis pigmentosa.

Ohlemiller *et al.* (1995) Neuroreport **6**:845-9 and Heckenlively *et al.* (1995) P.N.A.S. **92**:11100-11104 describe hearing loss and progressive retinal degeneration in tubby mice. The retinal degeneration is characterized by loss of photoreceptor cells, resulting in abnormal electroencephalograms by 3 weeks of age. Jones *et al.* (1992) Genomics **14**:197-9 localize the *tub* locus to a specific region of

chromosome 7, and demonstrate that it is distinct from the insulin-2 locus. The cholecystokinin receptor gene is shown to tightly linked to the *tub* locus in Samuelson *et al.* (1995) Genome 6:242-6. The mouse *tub* mutation is described in Coleman and Eicher (1990) J Hered 81:424-7 as an autosomal recessive mutation located on chromosome 7, which causes slowly developing but ultimately severe obesity.

Bennett *et al.* (1996) Nature Medicine 2:649 demonstrate that injection into *rd/rd* mice of a recombinant replication defective adenovirus that contains wild-type cDNA encoding βPDE delays photoreceptor death. Adenovirus vectors are described in Englehardt *et al.* (1993) Nature Genetics 4:27-34, and in Wang and Finer (1996) Nature Medicine 2:714.

Summary of the Invention

Nucleic acid compositions are provided that encode a family of mammalian proteins expressed in the retina and brain. Members of the gene family are genetically linked to various neurosensory defects, including cochlear degeneration, peripheral retinal degeneration and cone-rod retinal dystrophy. The nucleic acid compositions find use in identifying DNA sequences encoding homologous or related proteins; for production of the encoded protein; and in studying associated physiological pathways. In addition, modulation of the gene activity *in vivo* is used for prophylactic and therapeutic purposes, such as treatment of neurosensory defects, identification of retinal cells based on expression, and the like. The DNA is further used as a diagnostic for genetic predisposition to the linked neurosensory defect. One family member, *tub*, is associated with mature onset obesity in an animal model, and may be used as in assays and therapies directed to preventing or treating obesity.

Brief Description of the Drawings

Figure 1 illustrates the N-terminal splicing at the human and mouse *TUB* locus.

Figure 2A and Figure 2B show the intron/exon boundaries for TULP1 [SEQ ID NO:12] and TULP2 [SEQ ID NO:14]. The arrows above the sequence lines indicate splice junctions.

Description of the Specific Embodiments

A family of genes whose members are associated with various defects in sensory neurons are provided (TULP family). Among the linked diseases are cochlear defects, retinitis pigmentosa (RP-14) and combined rod-cone dystrophy (CRD). One family member, *tub*, is also associated with a genetic predisposition to adult onset obesity. The nucleotide sequences of human and mouse cDNAs and genomic regions are provided. The coding region sequences are highly conserved between family members at the carboxy terminus, and variable at the amino terminus.

The nucleic acid compositions find use in identifying DNA sequences encoding homologous or related proteins; for production of the encoded protein; and in studying associated physiological pathways *in vivo* and *in vitro*. The nucleic acids are useful in modulating gene activity for diagnostic, prophylactic and therapeutic purposes, such as treatment of neurosensory defects, identification of retinal cells based on expression, and the like. The DNA is further used as a diagnostic for genetic predisposition to the specific genetically linked defect.

The encoded proteins are useful as an immunogen to raise antibodies that specifically identify TULP expressing cells, in drug screening assays directed at neurosensory defects, and for therapeutic purposes. The amino terminal domain of TUB [SEQ ID NO:10, positions 1-139] has been shown to direct nuclear localization of the protein.

As used herein, the generic term "TULP" or "TULP family" designates the family of genes that includes the specific sequences provided in the SEQLIST and designated in Table 1. By family is intended one or more of the gene or gene products, up to and including TUB, TULP1, TULP2, TULP3 and TULP4. A family member is any one of the genes in the TULP family. Unless otherwise indicated, the sequences are of mammalian origin, and generally refer to the human sequences. In some animal models for TULP function, non-mammalian homologs, *e.g. C. elegans, D. melanogaster, etc.* are of interest. Within a species, the sequence similarity between family members is high in the carboxy terminal portion of the protein, where there is usually at least about 50% identity at the amino acid level. In *tub* and *tulp4* different transcriptional products are formed by alternative exon splicing in the 5' end of the gene. All members of the TULP family are expressed in the retina, although not for all splice variants. In some cases the genes are also expressed in other tissues.

Exemplary members of the TULP gene family are as follows:

TABLE 1

TULP FAMILY MEMBERS

SEQ ID NO	Sequence	Molecule	Size
1	Mouse *tub* Form I cDNA	dsDNA	2119 bp
2	translation of above	amino acid	459 aa
3	Mouse *tub* Form II cDNA	dsDNA	2434 bp
4	translation of above	amino acid	505 aa
5	*tub* mutation	dsDNA	480 bp
6	translation of above	amino acid	33 aa
7	Human *TUB* Form 6 cDNA	dsDNA	1426 bp
8	translation of above	amino acid	460 aa
9	Human *TUB* Form 1 cDNA	ds DNA	3060 bp
10	translation of above	amino acid	561 aa
11	Human *TUB* 5' region	genomic DNA	5995 bp
12	Human *TULP1* cDNA	ds DNA	2115 bp
13	translation of above	amino acid	542 aa
14	Human *TULP2* cDNA	ds DNA	1734 bp
15	translation of above	amino acid	520 aa
16	Human *TULP3* cDNA	ds DNA	1482 bp
17	translation of above	amino acid	442 aa
18	Mouse *TULP4* cDNA	ds DNA	1743 bp
19	translation of above	amino acid	506 aa
56	Human *TUB* Form 1; 5' RACE	ds cDNA	2112 bp
57	Human *TUB* Form 2; 5' RACE	ds cDNA	2368 bp
58	translation of above	amino acid	518 aa
59	Human *TUB* Form 3; 5' RACE	ds cDNA	1936 bp
60	translation of above	amino acid	512 aa
61	Human *TUB* Form 4; 5' RACE	ds cDNA	1890 bp
62	translation of above	amino acid	506 aa
63	Human *TUB* Form 5; 5' RACE	ds cDNA	2109 bp
64	Human *TUB* Form 6; 5' RACE	ds cDNA	2088 bp

The sequences of the human and mouse *tub* cDNA and encoded protein sequences are provided as SEQ ID NO:1 through 10. The genomic region 5' to the human *TUB* locus is provided as SEQ ID NO:11. The cDNA and encoded protein sequences of splicing variants of the human *TUB* locus are provided as SEQ ID Nos:56 through 64. Six cDNA splice variants of *TUB* have been identified, and are designated as Form 1 through 6. The encoded proteins have a common carboxy-terminal sequence [SEQ ID NO:8], and vary in the amino terminal sequences. Forms 1 through 4 have unique amino termini; Forms 5 and 6 vary from each other only in the non-translated cDNA sequences.

As used herein, *tub* designates a coding region, gene or gene product that maps to the exact chromosomal position of the *tub* mutation described by Coleman and Eicher, *supra*, and mammalian, particularly human, homologs thereof. The human *tub* locus maps to chromosome 11,

between the polymorphic markers D11S909 and D11S1331. It is expressed at high levels in brain, eye and testis, and at lower levels in various adult and fetal tissues, including small and large intestine, ovary and adipose tissue. Different transcriptional products are formed by alternative exon splicing in the 5' end of the gene.

The term "*tub*" or "*tubby*" encompasses both the normal mammalian sequence and the mutated sequence responsible for the *tub* phenotype. The *tub* mutation confers a genetic predisposition to maturity onset obesity in mice. The *tub* mutation is also associated with adult-onset degeneration of the retina and cochlea. The mutation in *tub/tub* mice is a G to T transversion at position 1704 resulting in a splicing defect and a truncated protein.

The sequence of the human *TULP1* gene and its predicted protein product are provided as SEQ ID NOs:12-13. The *TULP1* locus is associated with a predisposition to retinitis pigmentosa, form RP-14. *TULP1* localizes to human chromosome 6p21. Two markers, D6S439 and D6S291, that flank *TULP1* have been reported not to recombine with the RP 14 locus in a human kindred (Shugart *et al.* (1995) Am J Hum Genet. **57**:499-502) demonstrating that *TULP1* is tightly linked to the RP 14 locus. The expression of *TULP1* is restricted to the retina.

Loss of function mutations in *TULP1* have been shown to co-segregate with retinitis pigmentosa in kindred studies. Such mutations include but are not limited to a point mutation in exon 11 causing an amino acid substitution of Arg to Pro at a.a. 420 [SEQ ID NO:13]; and a point mutation in exon 12 causing an amino acid substitution of Phe to Leu at A.A 491 [SEQ ID NO:13]. The presently known polymorphisms that are associated with blindness are located in the conserved carboxy terminal portion of the protein.

The sequence of the human *TULP2* gene and its predicted protein product are provided as SEQ ID Nos:14-15. The expression of *TULP2* is restricted to the retina and testes. Retinal expression in adult tissue is relatively low. The *TULP2* locus is associated with a genetic predisposition to combined rod cone dystrophy, a disease causing early chorioretinal atrophy of the central and peripheral retina. *TULP2* is tightly linked to framework marker WI-9028 on chromosome 19q, which maps within the reported linked interval for CRD. The locus for rod cone dystrophy maps between D19S212 and D19S214.

The sequence of human *TULP3* and its predicted protein product are provided as SEQ ID Nos:16-17. The human *TULP3* gene maps to chromosome 12p13.2-12p13.3. The gene is expressed in the retina.

The sequence of mouse *tulp4* and its predicted protein product are provided as SEQ ID Nos:18-19. Different transcriptional products are formed by alternative exon splicing in the 5' end of the gene. The syntenic location of *TULP4* on the human chromosome is 19q.

TULP Nucleic Acid Compositions

Nucleic acids encoding TULP proteins may be cDNA, mRNA or genomic DNA, or a fragment thereof. The term "gene" shall be intended to mean an open reading frame encoding a specific TULP polypeptide, as exemplified in Table 1, as well as trancribed adjacent 5' and 3' non-coding nucleotide sequences, in either direction. The gene may further encompass non-transcribed regulatory regions adjacent to the transcribed regions. The gene may be introduced into an appropriate vector for extrachromosomal maintenance or for integration into the host.

The term "cDNA" as used herein is intended to include all nucleic acids that share the arrangement of sequence elements found in native mature mRNA species, where sequence elements are exons, 5' non-coding regions and 3' non-coding regions. Normally mRNA species have contiguous exons, with the intervening introns deleted, to create a continuous open reading frame.

Genomic *TULP* sequences have non-contiguous open reading frames, where introns interrupt the coding regions. A genomic sequence of interest comprises the nucleic acid present between an initiation codon and stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It may further include the 3' and 5' untranslated regions found in the mature mRNA. It may further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, *etc.*, including about 5 kb of flanking genomic DNA at either the 5' or 3' end of the coding region. The genomic DNA may be isolated as a fragment of 50 kbp or smaller. A preferred genomic sequence will lack those sequences that are linked to *TULP* in a native chromosome but which do not contribute to the biological function of the *TULP* gene.

Genomic regions of interest include the non-transcribed sequences 5' to a TULP family gene, usually from about one to six thousand bp of sequence. This region of DNA contains the native promoter elements that direct expression of the linked *TULP* gene. The non-transcribed region 5' to human *TUB* locus is provided in SEQ ID NO:11. The 3' portion of this sequence [nt. 5535 to 5995; SEQ ID NO:11] is

transcribed, but untranslated. The sequence of this 5' region may be utilized for promoter elements, including enhancer binding sites, that provide for expression in tissues where *TUB* is expressed. The tissue specific expression is useful for determining the pattern of expression, and for providing promoters that mimic the native pattern of expression. Methods for the identification of specific DNA motifs involved in the binding of transcriptional factors are known in the art, *e.g.* sequence similarity to known binding motifs, gel retardation studies, *etc.* For examples, see Blackwell *et al.* (1995) Mol Med **1**: 194-205; Mortlock *et al.* (1996) Genome Res. **6**: 327-33; and Joulin and Richard-Foy (1995) Eur J Biochem **232**: 620-626.

The nucleic acid compositions of the subject invention encode all or a part of the subject polypeptides. Fragments may be obtained of the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, by restriction enzyme digestion, by PCR amplification, *etc.* For the most part, DNA fragments will be of at least 25 nt, usually at least 30 nt, more usually at least about 50 nt. Such small DNA fragments are useful as primers for PCR, hybridization screening, *etc.* Larger DNA fragments, *i.e.* greater than 100 nt are useful for production of fragments of the encoded polypeptide.

Where it is desirable to generate probes or primers that distinguish one family member from other members of the gene family, sequences may be derived from the less conserved region of the genes. Such sequences include the 3' terminus, of about one thousand bp., of each of the TULP family cDNA sequences. Probes useful for identifying homologous genes, or multiple family members may be derived from the conserved region of the genes, which includes roughly the 5' 500-1000 bp of each of the TULP family cDNA sequences.

For use in amplification reactions, such as PCR, a pair of primers will be used. The exact composition of the primer sequences is not critical to the invention, but for most applications the primers will hybridize to the subject sequence under stringent conditions, as known in the art. It is preferable to choose a pair of primers that will generate an amplification product of at least about 50 nt, preferably at least about 100 nt. Algorithms for the selection of primer sequences are generally known, and are available in commercial software packages. Amplification primers hybridize to complementary strands of DNA, and will prime towards each other.

The DNA sequences are obtained in substantial purity, generally as a sequence other than a sequence of an intact mammalian chromosome. Usually, the DNA will be obtained substantially free of other nucleic acid sequences that do not include a *TULP* sequence or fragment thereof, generally being at least about 50%, usually at least about 90% pure and are typically "recombinant", *i.e.* flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

The DNA sequences may be used in a variety of ways. They may be used as probes for identifying other *TULP* genes, including novel family members, homologs and syntenic homologs. Identification of *TULP* homologs is based on similarity of sequence, chromosomal synteny, or both. The term homology is used to indicate a likeness of structure and conservation of biological function. Calculations of nucleic acid or amino acid sequence identity, as described below, provide a convenient method of identifying homologous or related genes, herein "homologs". Such homologs may be members of a gene family present in the same genome, or may be corresponding genes from different species. Chromosomal synteny may be used to further distinguish between homologous genes when there is sufficient evolutionary conservation between the genomes that are being compared, *e.g.* between mammalian species. A "syntenic homolog" has both sequence identity to the reference gene, and has the corresponding chromosomal location in relation to closely linked genes. Syntenic homologs have a high probability of sharing spatial and temporal localization of gene expression, and of encoding proteins that fill equivalent biological roles.

Mammalian homologs have substantial sequence similarity to the subject sequences, *i.e.* greater than 50% sequence identity with the amino acid or nucleotide sequence of the subject TULP sequence, as listed in Table 1. Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, *etc.* A reference sequence will usually be at least about 18 nt long, more usually at least about 30 nt long, and may extend to the complete sequence that is being compared. Algorithims for sequence analysis are known in the art, such as BLAST, described in Altschul *et al.* (1990) J Mol Biol 215:403-10.

Non-identical nucleic acids with sequence similarity are detected by hybridization under low stringency conditions, for example, at 50°C

and 10XSSC (0.9 M saline/0.09 M sodium citrate) and remain bound when subjected to washing at 55°C in 1XSSC. By using probes, particularly labeled probes of DNA sequences, one can isolate homologous or related genes. The source of homologous genes may be any mammalian species, *e.g.* primate species, particularly human; murines, such as rats and mice, canines, felines, bovines, ovines, equines, *etc.*

For hybridization probes, it may be desirable to use nucleic acid analogs, in order to improve the stability and and binding affinity. A number of modifications have been described that alter the chemistry of the phosphodiester backbone, sugars or heterocyclic bases.

Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O-phosphorothioate, 3'-CH2-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire phosphodiester backbone with a peptide linkage.

Sugar modifications are also used to enhance stability and affinity. The a-anomer of deoxyribose may be used, where the base is inverted with respect to the natural b-anomer. The 2'-OH of the ribose sugar may be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'-deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5- propynyl-2'-deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

Nucleic acid probes may also be used to identify expression of the gene in a biological specimen, *e.g.* retinal cells. The manner in which one probes cells for the presence of particular nucleotide sequences, as genomic DNA or RNA, is well-established in the literature and does not require elaboration here. A biological specimen is used as a source of mRNA. The mRNA may be amplified by RT-PCR, using reverse transcriptase to form a complementary DNA strand, followed by polymerase chain reaction amplification using primers specific for the subject DNA sequences. Alternatively, the mRNA sample is fractionated by electrophoresis, *e.g.* capillary or gel electrophoresis, transferred

to a suitable support, *e.g.* nitrocellulose and then probed with a fragment of the subject DNA as a probe. Other techniques may also find use, including oligonucleotide ligation assays, binding to solid state arrays, *etc.* Detection of mRNA having the subject sequence is indicative of *TULP* gene expression in the sample.

It will be understood by one of skill in the art that low basal levels of transcription are present in many normal cell types, or that a relatively rare cell type may have a high level of expression that cannot readily be detected in mRNA prepared from whole tissue. By specific expression, it is intended that mRNA levels are increased above the basal levels observed in other cells by at least about 100 fold, more usually by at least about 1000 fold. It will be further understood that malignant, or transformed, cells may express genes in an aberrant fashion.

Synthesis of TULP Proteins

The subject genes may be employed for producing all or portions of the TULP proteins. For expression, an expression cassette may be employed, providing for a transcriptional and translational initiation region, which may be inducible or constitutive, where the coding region is under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination region. Various transcriptional initiation regions may be employed which are functional in the expression host. In some cases, *e.g.* gene therapy vectors, it may be desirable to utilize the native promoter sequences as described above.

The peptide may be expressed in prokaryotes or eukaryotes in accordance with conventional ways, depending upon the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli, B. subtilis, S. cerevisiae,* or cells of a higher organism such as vertebrates, particularly mammals, *e.g.* COS 7 cells, may be used as the expression host cells. In many situations, it may be desirable to express the gene in mammalian cells, where the protein will benefit from native folding and post-translational modifications. Small peptides can also be synthesized in the laboratory.

With the availability of the protein in large amounts, by employing an expression host, the protein may be isolated and purified in accordance with conventional ways. A lysate may be prepared of the expression host and the lysate purified using HPLC, exclusion

chromatography, gel electrophoresis, affinity chromatography, or other purification technique. The purified protein will generally be at least about 80% pure, preferably at least about 90% pure, and may be up to and including 100% pure. Pure is intended to mean free of other proteins, as well as cellular debris.

A host may be treated with an intact TULP protein, or an active fragment thereof to modulate or reduce neurosensory and/or obesity-associated conditions. Desirably, the peptides will not induce an immune response, particularly an antibody response. Xenogeneic analogs may be screened for their ability to provide a therapeutic effect without raising an immune response. The protein or peptides may also be administered to *in vitro* cell cultures.

Various methods for administration may be employed. The polypeptide formulation may be given orally, or may be injected intravascularly, subcutaneously, peritoneally, *etc.* The dosage of the therapeutic formulation will vary widely, depending upon the nature of the disease, the frequency of administration, the manner of administration, the clearance of the agent from the host, and the like. The initial dose may be larger, followed by smaller maintenance doses. The dose may be administered as infrequently as weekly or biweekly, or fractionated into smaller doses and administered daily, semi-weekly, *etc.* to maintain an effective dosage level. In many cases, oral administration will require a higher dose than if administered intravenously. The amide bonds, as well as the amino and carboxy termini, may be modified for greater stability on oral administration.

The subject peptides may be prepared as formulations at a pharmacologically effective dose in pharmaceutically acceptable media, for example normal saline, PBS, *etc.* The additives may include bactericidal agents, stabilizers, buffers, or the like. In order to enhance the half-life of the subject peptide or subject peptide conjugates, the peptides may be encapsulated, introduced into the lumen of liposomes, prepared as a colloid, or another conventional technique may be employed that provides for an extended lifetime of the peptides.

The peptides may be administered as a combination therapy with other pharmacologically active agents. The additional drugs may be administered separately or in conjunction with the peptide compositions, and may be included in the same formulation.

The polypeptide is used for the production of antibodies, where short fragments provide for antibodies specific for the particular motif, and larger fragments or the entire protein allow for the

production of antibodies over the surface of the polypeptide. Antibodies may be raised to the wild-type or variant forms of TULP protein. Antibodies may be raised to isolated peptides corresponding to these domains, or to the native protein, *e.g.* by immunization with cells expressing a TULP gene, immunization with liposomes having a TULP protein inserted in the membrane, *etc.*

Antibodies are prepared in accordance with conventional ways, where the expressed polypeptide or protein is used as an immunogen, by itself or conjugated to known immunogenic carriers, *e.g.* KLH, pre-S HBsAg, other viral or eukaryotic proteins, or the like. Various adjuvants may be employed, with a series of injections, as appropriate. For monoclonal antibodies, after one or more booster injections, the spleen is isolated, the lymphocytes immortalized by cell fusion, and then screened for high affinity antibody binding. The immortalized cells, *i.e.* hybridomas, producing the desired antibodies may then be expanded. For further description, see Monoclonal Antibodies: A Laboratory Manual, Harlow and Lane eds., Cold Spring Harbor Laboratories, Cold Spring Harbor, New York, 1988. If desired, the mRNA encoding the heavy and light chains may be isolated and mutagenized by cloning in *E. coli*, and the heavy and light chains mixed to further enhance the affinity of the antibody. Alternatives to *in vivo* immunization as a method of raising antibodies include binding to phage "display" libraries, usually in conjunction with *in vitro* affinity maturation.

Diagnostic Uses

The subject compositions have a number of diagnostic uses, either as isolated family members, or as a panel of different sequences. The *TULP* genes and fragments thereof, encoded protein, and anti-TULP antibodies are useful in the identification of individuals predisposed to neurosensory degenerative conditions, *e.g.* cochlear degeneration and hearing loss; retinitis pigmentosa; combined rod cone dystrophy, *etc.* The characterization is useful in determining further treatment of the patient. Sequences of interest for diagnostic purposes include but are not limited to the conserved portion of the molecule as previously described. The conserved regions are identified by sequence similarity, and conservation of intron/exon structure.

Specifically, *TULP1* is associated with peripheral retinal dystrophies. In humans, *TULP1* is tightly linked to the RP-14 locus. *TUB* is associated with retinal degeneration and cochlear degeneration